

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
13 December 2001 (13.12.2001)

PCT

(10) International Publication Number  
**WO 01/94391 A2**

(51) International Patent Classification<sup>7</sup>: C07K 14/00

(21) International Application Number: PCT/US01/18595

(22) International Filing Date: 7 June 2001 (07.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/210,582 8 June 2000 (08.06.2000) US  
60/212,443 16 June 2000 (16.06.2000) US

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). HE, Ann [CN/US]; 4601 Catalina Drive, San Jose, CA 95129 (US). NGUYEN, Dannie, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). YAO, Monique, G. [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). BURFORD, Neil [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US).

HAFALIA, April [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). AZIMZAI, Yalda [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). WALLA, Narinder, K. [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be published upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/94391 A2

(54) Title: INTRACELLULAR SIGNALING PROTEINS

(57) Abstract: The invention provides human intracellular signaling proteins (ISIGP) and polynucleotides which identify and encode ISIGP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of ISIGP.

## INTRACELLULAR SIGNALING PROTEINS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of intracellular signaling proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, autoimmune/inflammatory, gastrointestinal, reproductive, and developmental disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of intracellular signaling proteins.

### BACKGROUND OF THE INVENTION

Intracellular signaling is the general process by which cells respond to extracellular signals (hormones, neurotransmitters, growth and differentiation factors, etc.) through a cascade of biochemical reactions that begins with the binding of a signaling molecule to a cell membrane receptor and ends with the activation of an intracellular target molecule. Intermediate steps in the process involve the activation of various cytoplasmic proteins by phosphorylation via protein kinases, and their deactivation by protein phosphatases, and the eventual translocation of some of these activated proteins to the cell nucleus where the transcription of specific genes is triggered. The intracellular signaling process regulates all types of cell functions including cell proliferation, cell differentiation, and gene transcription, and involves a diversity of molecules including protein kinases and phosphatases, and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens that regulate protein phosphorylation.

Certain proteins in intracellular signaling pathways serve to link or cluster other proteins involved in the signaling cascade. These proteins are referred to as scaffold, anchoring, or adaptor proteins. (For review, see Pawson, T., and Scott, J.D. (1997) *Science* 278:2075-2080.) As many intracellular signaling proteins such as protein kinases and phosphatases have relatively broad substrate specificities, the adaptors help to organize the component signaling proteins into specific biochemical pathways. Many of the above signaling molecules are characterized by the presence of particular domains that promote protein-protein interactions. A sampling of these domains is discussed below, along with other important intracellular messengers.

#### **Intracellular Signaling Second Messenger Molecules**

##### **Phospholipid and Inositol-phosphate Signaling**

Inositol phospholipids (phosphoinositides) are involved in an intracellular signaling pathway that begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane.

This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane to the biphosphate state ( $\text{PIP}_2$ ) by inositol kinases. Simultaneously, the G-protein linked receptor binding stimulates a trimeric G-protein which in turn activates a phosphoinositide-specific phospholipase C- $\beta$ . Phospholipase C- $\beta$  then cleaves  $\text{PIP}_2$  into two products, inositol triphosphate ( $\text{IP}_3$ ) and diacylglycerol. These two products act as mediators for separate signaling events.  $\text{IP}_3$  diffuses through the plasma membrane to induce calcium release from the endoplasmic reticulum (ER), while diacylglycerol remains in the membrane and helps activate protein kinase C, an STK that phosphorylates selected proteins in the target cell. The calcium response initiated by  $\text{IP}_3$  is terminated by the dephosphorylation of  $\text{IP}_3$  by specific inositol phosphatases. Cellular responses that are mediated by this pathway are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

#### Cyclic Nucleotide Signaling

Cyclic nucleotides (cAMP and cGMP) function as intracellular second messengers to transduce a variety of extracellular signals including hormones, light, and neurotransmitters. In particular, cyclic-AMP dependent protein kinases (PKA) are thought to account for all of the effects of cAMP in most mammalian cells, including various hormone-induced cellular responses. Visual excitation and the phototransmission of light signals in the eye is controlled by cyclic-GMP regulated,  $\text{Ca}^{2+}$ -specific channels. Because of the importance of cellular levels of cyclic nucleotides in mediating these various responses, regulating the synthesis and breakdown of cyclic nucleotides is an important matter. Thus adenylyl cyclase, which synthesizes cAMP from AMP, is activated to increase cAMP levels in muscle by binding of adrenaline to  $\beta$ -adrenergic receptors, while activation of guanylate cyclase and increased cGMP levels in photoreceptors leads to reopening of the  $\text{Ca}^{2+}$ -specific channels and recovery of the dark state in the eye. In contrast, hydrolysis of cyclic nucleotides by cAMP and cGMP-specific phosphodiesterases (PDEs) produces the opposite of these and other effects mediated by increased cyclic nucleotide levels. PDEs appear to be particularly important in the regulation of cyclic nucleotides, considering the diversity found in this family of proteins. At least seven families of mammalian PDEs (PDE1-7) have been identified based on substrate specificity and affinity, sensitivity to cofactors, and sensitivity to inhibitory drugs (Beavo, J.A. (1995) *Physiological Reviews* 75:725-48). PDE inhibitors have been found to be particularly useful in treating various clinical disorders. Rolipram, a specific inhibitor of PDE4, has been used in the treatment of depression, and similar inhibitors are undergoing evaluation as anti-inflammatory agents. Theophylline is a nonspecific PDE inhibitor used in the treatment of bronchial asthma and other respiratory diseases (Banner, K.H. and Page, C.P. (1995) *Eur. Respir. J.* 8:996-1000).

#### Calcium Signaling Molecules

Ca<sup>2+</sup> is another second messenger molecule that is even more widely used as an intracellular mediator than cAMP. Two pathways exist by which Ca<sup>2+</sup> can enter the cytosol in response to extracellular signals: One pathway acts primarily in nerve signal transduction where Ca<sup>2+</sup> enters a nerve terminal through a voltage-gated Ca<sup>2+</sup> channel. The second is a more ubiquitous pathway in which Ca<sup>2+</sup> is released from the ER into the cytosol in response to binding of an extracellular signaling molecule to a receptor. Ca<sup>2+</sup> directly activates regulatory enzymes, such as protein kinase C, which trigger signal transduction pathways. Ca<sup>2+</sup> also binds to specific Ca<sup>2+</sup>-binding proteins (CBPs) such as calmodulin (CaM) which then activate multiple target proteins in the cell including enzymes, membrane transport pumps, and ion channels. CaM interactions are involved in a multitude of cellular processes including, but not limited to, gene regulation, DNA synthesis, cell cycle progression, mitosis, cytokinesis, cytoskeletal organization, muscle contraction, signal transduction, ion homeostasis, exocytosis, and metabolic regulation (Celio, M.R. et al. (1996) Guidebook to Calcium-binding Proteins, Oxford University Press, Oxford, UK, pp. 15-20). Some Ca<sup>2+</sup> binding proteins are characterized by the presence of one or more EF-hand Ca<sup>2+</sup> binding motifs, which are comprised of 12 amino acids flanked by  $\alpha$ -helices (Celio, supra). The regulation of CBPs has implications for the control of a variety of disorders. Calcineurin, a CaM-regulated protein phosphatase, is a target for inhibition by the immunosuppressive agents cyclosporin and FK506. This indicates the importance of calcineurin and CaM in the immune response and immune disorders (Schwaninger M. et al. (1993) *J. Biol Chem.* 268:23111-23115). The level of CaM is increased several-fold in tumors and tumor-derived cell lines for various types of cancer (Rasmussen, C.D. and Means, A.R. (1989) *Trends in Neuroscience* 12:433-438).

### Signaling Complex Protein Domains

PDZ domains were named for three proteins in which this domain was initially discovered. These proteins include PSD-95 (postsynaptic density 95), Dlg (*Drosophila* lethal(1)discs large-1), and ZO-1 (zonula occludens-1). These proteins play important roles in neuronal synaptic transmission, tumor suppression, and cell junction formation, respectively. Since the discovery of these proteins, over sixty additional PDZ-containing proteins have been identified in diverse prokaryotic and eukaryotic organisms. This domain has been implicated in receptor and ion channel clustering and in the targeting of multiprotein signaling complexes to specialized functional regions of the cytosolic face of the plasma membrane. (For review of PDZ domain-containing proteins, see Ponting, C. P. et al. (1997) *Bioessays* 19:469-479.) A large proportion of PDZ domains are found in the eukaryotic MAGUK (membrane-associated guanylate kinase) protein family, members of which bind to the intracellular domains of receptors and channels. However, PDZ domains are also found in diverse membrane-localized proteins

such as protein tyrosine phosphatases, serine/threonine kinases, G-protein cofactors, and synapse-associated proteins such as syntrophins and neuronal nitric oxide synthase (nNOS). Generally, about one to three PDZ domains are found in a given protein, although up to nine PDZ domains have been identified in a single protein. The glutamate receptor interacting protein (GRIP) contains seven PDZ domains. GRIP is an adaptor that links certain glutamate receptors to other proteins and may be responsible for the clustering of these receptors at excitatory synapses in the brain (Dong, H. et al. (1997) *Nature* 386:279-284).

The SH3 domain is defined by homology to a region of the proto-oncogene c-Src, a cytoplasmic protein tyrosine kinase. SH3 is a small domain of 50 to 60 amino acids that interacts with proline-rich ligands. SH3 domains are found in a variety of eukaryotic proteins involved in signal transduction, cell polarization, and membrane-cytoskeleton interactions. In some cases, SH3 domain-containing proteins interact directly with receptor tyrosine kinases. For example, the SLAP-130 protein is a substrate of the T-cell receptor (TCR) stimulated protein kinase. SLAP-130 interacts via its SH3 domain with the protein SLP-76 to affect the TCR-induced expression of interleukin-2 (Musci, M.A. et al. (1997) *J. Biol. Chem.* 272:11674-11677). Another recently identified SH3 domain protein is macrophage actin-associated tyrosine-phosphorylated protein (MAYP) which is phosphorylated during the response of macrophages to colony stimulating factor-1 (CSF-1) and is likely to play a role in regulating the CSF-1-induced reorganization of the actin cytoskeleton (Yeung, Y.-G. et al. (1998) *J. Biol. Chem.* 273:30638-30642). The structure of SH3 is characterized by two antiparallel beta sheets packed against each other at right angles. This packing forms a hydrophobic pocket lined with residues that are highly conserved between different SH3 domains. This pocket makes critical hydrophobic contacts with proline residues in the ligand (Feng, S. et al. (1994) *Science* 266: 1241-47).

The pleckstrin homology (PH) domain was originally identified in pleckstrin, the predominant substrate for protein kinase C in platelets. Since its discovery, this domain has been identified in over 90 proteins involved in intracellular signaling or cytoskeletal organization. Proteins containing the pleckstrin homology domain include a variety of kinases, phospholipase-C isoforms, guanine nucleotide release factors, and GTPase activating proteins. For example, members of the FGD1 family contain both Rho-guanine nucleotide exchange factor (GEF) and PH domains, as well as a FYVE zinc finger domain. FGD1 is the gene responsible for faciogenital dysplasia, an inherited skeletal dysplasia (Pasteris, N.G. and Gorski, J.L. (1999) *Genomics* 60:57-66). Many PH domain proteins function in association with the plasma membrane, and this association appears to be mediated by the PH domain itself. PH domains share a common structure composed of two antiparallel beta sheets flanked by an amphipathic alpha helix. Variable loops connecting the component beta strands generally occur within a positively charged environment and may function as ligand binding sites. (Lemmon, M. A. et al.

(1996) Cell 85:621-624.)

The tetratricopeptide repeat (TPR) is a 34 amino acid repeated motif found in organisms from bacteria to humans. TPRs are predicted to form amphipathic helices, and appear to mediate protein-protein interactions. TPR domains are found in CDC16, CDC23, and CDC27, members of the anaphase promoting complex which targets proteins for degradation at the onset of anaphase. Other processes involving TPR proteins include cell cycle control, transcription repression, stress response, and protein kinase inhibition. (Lamb, J.R. et al. (1995) Trends Biochem. Sci. 20:257-259.)

The armadillo/beta-catenin repeat is a 42 amino acid motif which forms a superhelix of alpha helices when tandemly repeated. The structure of the armadillo repeat region from beta-catenin revealed a shallow groove of positive charge on one face of the superhelix, which is a potential binding surface. The armadillo repeats of beta-catenin, plakoglobin, and p120<sup>cas</sup> bind the cytoplasmic domains of cadherins. Beta-catenin/cadherin complexes are targets of regulatory signals that govern cell adhesion and mobility. (Huber, A.H. et al. (1997) Cell 90:871-882.)

The WW domain binds to proline-rich ligands. The structure of the WW domain is composed of beta strands grouped around four conserved aromatic residues, generally tryptophan. This domain was originally discovered in dystrophin, a cytoskeletal protein with direct involvement in Duchenne muscular dystrophy (Bork, P. and M. Sudol (1994) Trends Biochem. Sci. 19:531-533). WW domains have since been discovered in a variety of intracellular signaling molecules involved in development, cell differentiation, and cell proliferation. Signaling complexes mediated by WW domains have been implicated in several human diseases, including Liddle's syndrome of hypertension, muscular dystrophy, and Alzheimer's disease (Sudol, *supra*).

ANK repeats mediate protein-protein interactions associated with diverse intracellular signaling functions. For example, ANK repeats are found in proteins involved in cell proliferation such as kinases, kinase inhibitors, tumor suppressors, and cell cycle control proteins. (See, for example, Kalus, W. et al. (1997) FEBS Lett. 401:127-132; Ferrante, A. W. et al. (1995) Proc. Natl. Acad. Sci. USA 92:1911-1915.) These proteins generally contain multiple ANK repeats, each composed of about 33 amino acids. Myotrophin is an ANK repeat protein that plays a key role in the development of cardiac hypertrophy, a contributing factor to many heart diseases. Structural studies show that the myotrophin ANK repeats, like other ANK repeats, each form a helix-turn-helix core preceded by a protruding "tip." These tips are of variable sequence and may play a role in protein-protein interactions. The helix-turn-helix region of the ANK repeats stack on top of one another and are stabilized by hydrophobic interactions (Yang, Y. et al. (1998) Structure 6:619-626). Another example of an ANK repeat protein is the C. elegans FEM1 protein and its mammalian homologs, which mediate apoptosis during development (Ventura-Holman, T. et al. (1998) Genomics 54:221-230).

The final step in cell signaling pathways is the transcription of specific genes, often mediated by the activation of selected transcriptional regulatory proteins. Some of these proteins function as transcription factors that initiate, activate, repress, or terminate gene transcription. Transcription factors generally bind to promoter, enhancer, or upstream regulatory regions of a gene in a

5 sequence-specific manner, although some factors bind regulatory elements within or downstream of the coding region. Transcription factors may bind to a specific region of DNA singly or as a complex with other accessory factors. (Reviewed in Lewin, B. (1990) Genes IV, Oxford University Press, New York, NY, pp. 554-570.)

The zinc finger motif, which binds zinc ions, generally contains tandem repeats of about 30  
10 amino acids consisting of periodically spaced cysteine and histidine residues. Examples of this sequence pattern include the C2H2-type, C4-type, and C3HC4-type zinc fingers, and the PHD domain (Lewin, supra; Aasland, R., et al. (1995) Trends Biochem. Sci. 20:56-59). Zinc finger proteins each contain an  $\alpha$  helix and an antiparallel  $\beta$  sheet whose proximity and conformation are maintained by the zinc ion. Contact with DNA is made by the arginine preceding the  $\alpha$  helix and by the second, third, and  
15 sixth residues of the  $\alpha$  helix.

Many neoplastic disorders in humans can be attributed to inappropriate gene expression. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes (Cleary, M.L. (1992) Cancer Surv. 15:89-104). One clinically relevant zinc-finger protein is WT1, a tumor-suppressor protein that is inactivated in children  
20 with Wilm's tumor. The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, A.G. (1995) N. Engl. J. Med. 332:45-47).

In addition, the immune system responds to infection or trauma by activating a cascade of events that coordinate the progressive selection, amplification, and mobilization of cellular defense mechanisms. A complex and balanced program of gene activation and repression is involved in this  
25 process. However, hyperactivity of the immune system as a result of improper or insufficient regulation of gene expression may result in considerable tissue or organ damage. This damage is well documented in immunological responses associated with arthritis, allergens, heart attack, stroke, and infections (Isselbacher et al. Harrison's Principles of Internal Medicine, 13/e, McGraw Hill, Inc. and Teton Data Systems Software, 1996). The causative gene for autoimmune polyendocrinopathy-candidiasis-  
30 ectodermal dystrophy (APECED) was recently isolated and found to encode a protein with two PHD-type zinc finger motifs (Bjorses, P. et al. (1998) Hum. Mol. Genet. 7:1547-1553).

Furthermore, the generation of multicellular organisms is based upon the induction and coordination of cell differentiation at the appropriate stages of development. Central to this process is differential gene expression, which confers the distinct identities of cells and tissues throughout the

body. Failure to regulate gene expression during development can result in developmental disorders. Zinc finger proteins involved in the determination of cell fate include *deltex*, a regulator of the notch receptor signaling pathway which regulates many cell fate decisions during development (Frolova, E. and Beebe, D. (2000) *Mech. Dev.* 92:285-289), and the recently isolated g1 related protein (G1RP),  
5 which appears to regulate growth factor withdrawal-induced apoptosis of myeloid precursor cells (Baker, S.J. and Reddy, E.P. (2000) *Gene* 248:33-40). Human developmental disorders caused by mutations in zinc finger-type transcriptional regulators include: urogenital developmental abnormalities associated with WT1; Greig cephalopolysyndactyly, Pallister-Hall syndrome, and postaxial polydactyly type A (GLI3); and Townes-Brocks syndrome, characterized by anal, renal, limb,  
10 and ear abnormalities (SALL1) (Engelkamp, D. and van Heyningen, V. (1996) *Curr. Opin. Genet. Dev.* 6:334-342; Kohlhase, J. et al. (1999) *Am. J. Hum. Genet.* 64:435-445).

The discovery of new intracellular signaling proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, gastrointestinal, reproductive, and  
15 developmental disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of intracellular signaling proteins.

#### SUMMARY OF THE INVENTION

The invention features purified polypeptides, intracellular signaling proteins, referred to  
20 collectively as "ISIGP" and individually as "ISIGP-1," "ISIGP-2," "ISIGP-3," "ISIGP-4," and "ISIGP-5." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a  
25 biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-5.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from  
30 the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an



amino acid sequence selected from the group consisting of SEQ ID NO:1-5. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-5. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:6-10.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter  
5 sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of  
10 SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group  
15 consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid  
20 sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a  
25 polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a  
30 polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, c) a

polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected

from the group consisting of SEQ ID NO:1-5. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional ISIGP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional ISIGP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional ISIGP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid

sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a  
5 compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from  
10 the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in  
15 the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering  
20 expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:6-10, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method  
25 comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a  
30 polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising

a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will

be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a  
5 reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to  
10 practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## 15 DEFINITIONS

"ISIGP" refers to the amino acid sequences of substantially purified ISIGP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of  
20 ISIGP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ISIGP either by directly interacting with ISIGP or by acting on components of the biological pathway in which ISIGP participates.

An "allelic variant" is an alternative form of the gene encoding ISIGP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in  
25 polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

30 "Altered" nucleic acid sequences encoding ISIGP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as ISIGP or a polypeptide with at least one functional characteristic of ISIGP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding ISIGP, and improper or unexpected hybridization to allelic variants, with a

locus other than the normal chromosomal locus for the polynucleotide sequence encoding ISIGP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent ISIGP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of ISIGP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of ISIGP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ISIGP either by directly interacting with ISIGP or by acting on components of the biological pathway in which ISIGP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind ISIGP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to

immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

5           The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having  
10 modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the  
15 designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

          The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic ISIGP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific  
20 antibodies.

          "Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

          A "composition comprising a given polynucleotide sequence" and a "composition comprising a  
25 given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding ISIGP or fragments of ISIGP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be  
30 deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

          "Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from



one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

5        "Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

10	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
15	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
20	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
25	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
30	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

35        A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological

function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

A "fragment" is a unique portion of ISIGP or the polynucleotide encoding ISIGP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:6-10 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:6-10, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:6-10 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:6-10 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:6-10 and the region of SEQ ID NO:6-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-5 is encoded by a fragment of SEQ ID NO:6-10. A fragment of SEQ ID NO:1-5 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-5. For example, a fragment of SEQ ID NO:1-5 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-5. The precise length of a fragment of SEQ ID NO:1-5 and the region of SEQ ID NO:1-5 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon

(e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

5       The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

10       Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191.

15       For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

20       Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at

<http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other  
25       polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to  
30       compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

5 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such  
10 lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in  
15 a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment  
20 methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment  
25 program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

30 Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

- 5       Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length  
10 supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

      "Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

- 15       The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

- "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific  
20 hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for  
25 annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

- 30       Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions

for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of ISIGP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of ISIGP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of ISIGP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of ISIGP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, 5 polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked 10 to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of 15 amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an ISIGP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the 20 art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of ISIGP.

"Probe" refers to nucleic acid sequences encoding ISIGP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels 25 include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

30 Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the

specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular  
5 Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR  
Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

10 Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU  
15 primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer  
20 binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that  
25 hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary  
30 polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the



artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, *supra*. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence.

5 Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated  
10 regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent,  
15 chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose  
20 instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing ISIGP, nucleic acids encoding ISIGP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

25 The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope  
30 A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which

they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least

60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human intracellular signaling proteins (ISIGP), the polynucleotides encoding ISIGP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, gastrointestinal, reproductive, and developmental disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by

BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog.

- 5 Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

- Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.
- 10
- 15

- Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are intracellular signaling proteins. For example, SEQ ID NO:1 is 36% identical to rat membrane-associated guanylate kinase-interacting protein (GenBank ID g4151807) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $5.0e-13$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:1 is a membrane-associated guanylate kinase-interacting protein. In an alternative example, SEQ ID NO:4 is 43% identical to mouse g1-related zinc finger protein (GenBank ID g6175860) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $5.0e-60$ . SEQ ID NO:4 also contains a zinc finger C3HC4 type (RING finger) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:4 is a zinc finger-type transcriptional regulator. SEQ ID NO:2, SEQ ID NO:3, and SEQ ID NO:5 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-5 are described in Table 7.
- 20
- 25
- 30

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:6-10 or that distinguish between SEQ ID NO:6-10 and related polynucleotide sequences. Column 3 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 4 and 5 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 3 relative to their respective full length sequences.

The identification numbers in Column 3 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 1617090F6 is the identification number of an Incyte cDNA sequence, and BRAITUT12 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70794548V1). Alternatively, the identification numbers in column 3 may refer to GenBank cDNAs or ESTs (e.g., g6140473) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 3 may refer to coding regions predicted by Genscan analysis of genomic DNA. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 3 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. (See Example V.) Alternatively, the identification numbers in column 3 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 3 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses ISIGP variants. A preferred ISIGP variant is one which has at

least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the ISIGP amino acid sequence, and which contains at least one functional or structural characteristic of ISIGP.

The invention also encompasses polynucleotides which encode ISIGP. In a particular  
5 embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:6-10, which encodes ISIGP. The polynucleotide sequences of SEQ ID NO:6-10, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

10 The invention also encompasses a variant of a polynucleotide sequence encoding ISIGP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding ISIGP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:6-10 which has at  
15 least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:6-10. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of ISIGP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic  
20 code, a multitude of polynucleotide sequences encoding ISIGP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally  
25 occurring ISIGP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode ISIGP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring ISIGP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding ISIGP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring  
30 codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding ISIGP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced

from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode ISIGP and ISIGP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using  
5 reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding ISIGP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:6-10 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and  
10 S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of  
15 DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV),  
20 PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols  
25 in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding ISIGP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed,  
30 restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids*

Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown  
5 sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using  
10 commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been  
15 size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the  
20 size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire  
25 process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode ISIGP may be cloned in recombinant DNA molecules that direct expression of ISIGP, or  
30 fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express ISIGP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter ISIGP-encoding sequences for a variety of purposes including, but not



limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of ISIGP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding ISIGP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, ISIGP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of ISIGP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing.

(See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active ISIGP, the nucleotide sequences encoding ISIGP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding ISIGP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding ISIGP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding ISIGP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding ISIGP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding ISIGP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and

Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding ISIGP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding ISIGP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding ISIGP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of ISIGP are needed, e.g. for the production of antibodies, vectors which direct high level expression of ISIGP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of ISIGP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of ISIGP. Transcription of sequences encoding ISIGP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp.

191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding ISIGP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses ISIGP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of ISIGP in cell lines is preferred. For example, sequences encoding ISIGP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *ap<sup>r</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used.

These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is  
5 also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding ISIGP is inserted within a marker gene sequence, transformed cells containing sequences encoding ISIGP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding ISIGP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates  
10 expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding ISIGP and that express ISIGP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based  
15 technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of ISIGP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal  
20 antibodies reactive to two non-interfering epitopes on ISIGP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) *Serological Methods, a Laboratory Manual*, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) *Current Protocols in Immunology*, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) *Immunochemical Protocols*, Humana Press, Totowa NJ.)

25 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding ISIGP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding ISIGP, or any fragments thereof, may be cloned into a vector for the production of  
30 an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include

radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding ISIGP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein  
5 produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode ISIGP may be designed to contain signal sequences which direct secretion of ISIGP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the  
10 inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities  
15 (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding ISIGP may be ligated to a heterologous sequence resulting in translation of a fusion  
20 protein in any of the aforementioned host systems. For example, a chimeric ISIGP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of ISIGP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose  
25 binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion  
30 proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the ISIGP encoding sequence and the heterologous protein sequence, so that ISIGP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled ISIGP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for  
5 example, <sup>35</sup>S-methionine.

ISIGP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to ISIGP. At least one and up to a plurality of test compounds may be screened for specific binding to ISIGP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

10 In one embodiment, the compound thus identified is closely related to the natural ligand of ISIGP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which ISIGP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the  
15 compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express ISIGP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing ISIGP or cell membrane fractions which contain ISIGP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either ISIGP or the compound is  
20 analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with ISIGP, either in solution or affixed to a solid support, and detecting the binding of ISIGP to the compound. Alternatively, the assay  
25 may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

ISIGP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of ISIGP. Such compounds may include agonists, antagonists, or partial or  
30 inverse agonists. In one embodiment, an assay is performed under conditions permissive for ISIGP activity, wherein ISIGP is combined with at least one test compound, and the activity of ISIGP in the presence of a test compound is compared with the activity of ISIGP in the absence of the test compound. A change in the activity of ISIGP in the presence of the test compound is indicative of a compound that modulates the activity of ISIGP. Alternatively, a test compound is combined with an in

vitro or cell-free system comprising ISIGP under conditions suitable for ISIGP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of ISIGP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

- 5 In another embodiment, polynucleotides encoding ISIGP or their mammalian homologs may be “knocked out” in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo  
10 and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D.  
15 (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.
- 20 Polynucleotides encoding ISIGP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).
- 25 Polynucleotides encoding ISIGP can also be used to create “knockin” humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding ISIGP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential  
30 pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress ISIGP, e.g., by secreting ISIGP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

#### THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between



regions of ISIGP and intracellular signaling proteins. In addition, the expression of ISIGP is closely associated with placenta tissue, neonatal keratinocytes, and prostate epithelial tissue. Therefore, ISIGP appears to play a role in cell proliferative, autoimmune/inflammatory, gastrointestinal, reproductive, and developmental disorders. In the treatment of disorders associated with increased  
 5 ISIGP expression or activity, it is desirable to decrease the expression or activity of ISIGP. In the treatment of disorders associated with decreased ISIGP expression or activity, it is desirable to increase the expression or activity of ISIGP.

Therefore, in one embodiment, ISIGP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ISIGP.

10 Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone,  
 15 bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia,  
 20 autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome,  
 25 multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and  
 30 helminthic infections, and trauma; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis,

- hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice,
- 5 hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease,  $\alpha_1$ -antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and a hepatic tumor including a nodular hyperplasia, an adenoma, and a carcinoma; a
- 10 reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology,
- 15 cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puberty, retrograde ejaculation and anejaculation, haemangioblastomas, cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac
- 20 tumours; and a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis,
- 25 hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss.

In another embodiment, a vector capable of expressing ISIGP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ISIGP including, but not limited to, those described above.

- 30 In a further embodiment, a composition comprising a substantially purified ISIGP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ISIGP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of ISIGP may be

administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ISIGP including, but not limited to, those listed above.

In a further embodiment, an antagonist of ISIGP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ISIGP. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, gastrointestinal, reproductive, and developmental disorders described above. In one aspect, an antibody which specifically binds ISIGP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express ISIGP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding ISIGP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ISIGP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of ISIGP may be produced using methods which are generally known in the art. In particular, purified ISIGP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind ISIGP. Antibodies to ISIGP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with ISIGP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to ISIGP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at

least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of ISIGP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

5            Monoclonal antibodies to ISIGP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and  
10    Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

          In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda,  
15    S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce ISIGP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

20            Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

          Antibody fragments which contain specific binding sites for ISIGP may also be generated. For  
25    example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

30            Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between ISIGP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two

non-interfering ISIGP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for ISIGP. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of ISIGP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple ISIGP epitopes, represents the average affinity, or avidity, of the antibodies for ISIGP. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular ISIGP epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the ISIGP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of ISIGP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of ISIGP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding ISIGP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding ISIGP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding ISIGP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998)

J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include

5 liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding ISIGP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency

10 (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene

15 Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g.,

20 against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in ISIGP expression or regulation causes disease, the expression of

25 ISIGP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in ISIGP are treated by constructing mammalian expression vectors encoding ISIGP and introducing these vectors by mechanical means into ISIGP-deficient cells. Mechanical transfer technologies for use with

30 cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of ISIGP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). ISIGP may be expressed

5 using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the

10 ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding ISIGP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID

15 TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these

20 standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to ISIGP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding ISIGP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive

25 element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for

30 receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a

method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference.

Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding ISIGP to cells which have one or more genetic abnormalities with respect to the expression of ISIGP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding ISIGP to target cells which have one or more genetic abnormalities with respect to the expression of ISIGP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing ISIGP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary



skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding ISIGP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for ISIGP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of ISIGP-coding RNAs and the synthesis of high levels of ISIGP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of ISIGP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding ISIGP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by

scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding ISIGP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding ISIGP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased ISIGP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding ISIGP may be therapeutically useful, and in the treatment of disorders associated with decreased ISIGP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding ISIGP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method

commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding ISIGP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding ISIGP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding ISIGP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various

formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of ISIGP, antibodies to ISIGP, and mimetics, agonists, antagonists, or inhibitors of ISIGP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising ISIGP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, ISIGP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example ISIGP or fragments thereof, antibodies of ISIGP, and agonists, antagonists or inhibitors of ISIGP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by

calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are  
5 used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject  
10 requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending  
15 on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their  
20 inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind ISIGP may be used for the diagnosis of disorders characterized by expression of ISIGP, or in assays to monitor patients being treated with  
25 ISIGP or agonists, antagonists, or inhibitors of ISIGP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for ISIGP include methods which utilize the antibody and a label to detect ISIGP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter  
30 molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring ISIGP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of ISIGP expression. Normal or standard values for ISIGP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to ISIGP under

conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of ISIGP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

5 In another embodiment of the invention, the polynucleotides encoding ISIGP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of ISIGP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of  
10 ISIGP, and to monitor regulation of ISIGP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding ISIGP or closely related molecules may be used to identify nucleic acid sequences which encode ISIGP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a  
15 conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding ISIGP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the ISIGP-encoding sequences. The hybridization probes of the subject  
20 invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:6-10 or from genomic sequences including promoters, enhancers, and introns of the ISIGP gene.

Means for producing specific hybridization probes for DNAs encoding ISIGP include the cloning of polynucleotide sequences encoding ISIGP or ISIGP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may  
25 be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding ISIGP may be used for the diagnosis of disorders associated  
30 with expression of ISIGP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in

particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's

5 disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis,

10 glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner

15 syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the

20 intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome

25 (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and a hepatic tumor including a nodular hyperplasia,

30 an adenoma, and a carcinoma; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a disruption of

spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puberty, retrograde ejaculation and

5 anejaculation, haemangioblastomas, cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumours; and a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome,

10 hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss. The polynucleotide sequences encoding ISIGP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR

15 technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered ISIGP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding ISIGP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide

20 sequences encoding ISIGP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding ISIGP in the sample

25 indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of ISIGP, a normal or standard profile for expression is established. This may be accomplished by combining

30 body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding ISIGP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from



patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding ISIGP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding ISIGP, or a fragment of a polynucleotide complementary to the polynucleotide encoding ISIGP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding ISIGP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding ISIGP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-

based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

5       Methods which may also be used to quantify the expression of ISIGP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of  
10 interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large  
15 numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used  
20 to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, ISIGP, fragments of ISIGP, or antibodies specific for ISIGP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein  
25 interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a  
30 given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present

invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global

pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and  
5 analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent  
10 such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for  
15 example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for ISIGP to quantify the  
20 levels of ISIGP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoze, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-  
25 reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the  
30 analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological

sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding ISIGP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map

data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding ISIGP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

5        In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery  
10 techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc.,  
15 among normal, carrier, or affected individuals.

In another embodiment of the invention, ISIGP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes  
20 between ISIGP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with ISIGP, or fragments thereof, and  
25 washed. Bound ISIGP is then detected by methods well known in the art. Purified ISIGP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding ISIGP specifically compete with a test compound for binding ISIGP. In  
30 this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with ISIGP.

In additional embodiments, the nucleotide sequences which encode ISIGP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such

properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/210,582 and U.S. Ser. No. 60/212,443, are expressly incorporated by reference herein.

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 3. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g.,

PBLUESCRIPT plasmid (Stratagene), PSPT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX

5 DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC  
10 Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a  
15 high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## 20 III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB  
25 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or  
30 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.



The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID

NO:6-10. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

#### IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative intracellular signaling proteins were initially identified by running the Genscan gene  
5 identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a  
general-purpose gene identification program which analyzes genomic DNA sequences from a variety of  
organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin  
(1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an  
assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a  
10 FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for  
Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA  
sequences encode intracellular signaling proteins, the encoded polypeptides were analyzed by querying  
against PFAM models for intracellular signaling proteins. Potential intracellular signaling proteins  
were also identified by homology to Incyte cDNA sequences that had been annotated as intracellular  
15 signaling proteins. These selected Genscan-predicted sequences were then compared by BLAST  
analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences  
were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence  
predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any  
Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence  
20 for transcription. When Incyte cDNA coverage was available, this information was used to correct or  
confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by  
assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA  
sequences using the assembly process described in Example III. Alternatively, full length  
polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding  
25 sequences.

#### V. Assembly of Genomic Sequence Data with cDNA Sequence Data

##### "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification  
30 program described in Example IV. Partial cDNAs assembled as described in Example III were mapped  
to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from  
one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory  
and dynamic programming to integrate cDNA and genomic information, generating possible splice  
variants that were subsequently confirmed, edited, or extended to create a full length sequence.

Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpr public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

#### "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### **VI. Chromosomal Mapping of ISIGP Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:6-10 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:6-10 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome

Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

#### VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

25

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The

product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is  
5 produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding ISIGP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the  
10 following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories.  
15 Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding ISIGP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ  
20 GOLD database (Incyte Genomics, Palo Alto CA).

#### VIII. Extension of ISIGP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was  
25 synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

30 Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ ,

and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the  
5 alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE  
10 and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

15 The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were  
20 religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

25 The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified  
30 using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or

are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:6-10 are employed to screen cDNAs, genomic  
5 DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston  
10 MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

15 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and  
20 compared.

#### X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned  
25 technologies should be uniform and solid with a non-porous surface (Schena (1999), *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may  
30 contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be

selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a  
5 fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### 10 Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand  
15 buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20  
20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended  
25 in 14  $\mu$ l 5X SSC/0.2% SDS.

#### Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified  
30 in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific



Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average  
5 concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate  
10 buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

#### Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample  
15 mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS),  
20 three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

#### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is  
25 focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially.  
30 Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source,

although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### **XI. Complementary Polynucleotides**

Sequences complementary to the ISIGP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring ISIGP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of ISIGP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the ISIGP-encoding transcript.

#### **XII. Expression of ISIGP**

Expression and purification of ISIGP is achieved using bacterial or virus-based expression systems. For expression of ISIGP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription.

Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element.

Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express ISIGP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG).

- 5 Expression of ISIGP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding ISIGP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of
- 10 cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

- In most expression systems, ISIGP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,
- 15 affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from ISIGP at
  - 20 specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified ISIGP obtained by these methods can be used directly in the assays shown in
  - 25 Examples XVI and XVII where applicable.

### XIII. Functional Assays

- ISIGP function is assessed by expressing the sequences encoding ISIGP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice
- 30 include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish

transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of ISIGP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding ISIGP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding ISIGP and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### XIV. Production of ISIGP Specific Antibodies

ISIGP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the ISIGP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the

oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-ISIGP activity by, for example, binding the peptide or ISIGP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### **XV. Purification of Naturally Occurring ISIGP Using Specific Antibodies**

5 Naturally occurring or recombinant ISIGP is substantially purified by immunoaffinity chromatography using antibodies specific for ISIGP. An immunoaffinity column is constructed by covalently coupling anti-ISIGP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

10 Media containing ISIGP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of ISIGP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/ISIGP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and ISIGP is collected.

#### **15 XVI. Identification of Molecules Which Interact with ISIGP**

ISIGP, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled ISIGP, washed, and any wells with labeled ISIGP complex are assayed. Data obtained using different concentrations of  
20 ISIGP are used to calculate values for the number, affinity, and association of ISIGP with the candidate molecules.

Alternatively, molecules interacting with ISIGP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

25 ISIGP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### **XVII. Demonstration of ISIGP Activity**

30 An assay for ISIGP activity is based on a prototypical assay for ligand/receptor-mediated modulation of cell proliferation. This assay measures the amount of newly synthesized DNA in Swiss mouse 3T3 cells expressing ISIGP. cDNA encoding ISIGP is subcloned into a mammalian expression vector that drives high levels of cDNA transcription. This recombinant vector is transfected into quiescent 3T3 cultured cells using methods well known in the art. The transfected cells are incubated in

the presence of [<sup>3</sup>H]thymidine. Incorporation of [<sup>3</sup>H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a tritium radioisotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. Statistically significant stimulation of DNA synthesis in the presence of the recombinant vector, relative to that in non-transfected cells, is indicative of ISIGP activity.

Alternatively, ISIGP activity is associated with its ability to form protein-protein complexes and is measured by its ability to regulate growth characteristics of NIH3T3 mouse fibroblast cells. A cDNA encoding ISIGP is subcloned into an appropriate eukaryotic expression vector. This vector is transfected into NIH3T3 cells using methods known in the art. Transfected cells are compared with non-transfected cells for the following quantifiable properties: growth in culture to high density, reduced attachment of cells to the substrate, altered cell morphology, and ability to induce tumors when injected into immunodeficient mice. The activity of ISIGP is proportional to the extent of increased growth or frequency of altered cell morphology in NIH3T3 cells transfected with ISIGP.

ISIGP-1 activity may be demonstrated by measuring the interaction of ISIGP-1 with a guanylate kinase such as synaptic scaffolding molecule (S-SCAM) (Yao, I. et al. (1999) *J. Biol. Chem.* 274:11889-11896). Samples of ISIGP-1 are fixed on glutathione-Sepharose 4B beads. COS cells are cultured with 10% fetal bovine serum under 10% CO<sub>2</sub> at 37 °C. Two 10cm plates of COS cells are homogenized in 0.5 ml of 20 mM Tris/HCl, pH 7.4, at 100,000g for 30 min. Aliquots of 0.5 ml COS cell extract are incubated with ISIGP-1 fixed on 20 µl glutathione beads. S-SCAM attached to the beads is detected by SDS-polyacrylamide gel electrophoresis and immunoblotting using, for example, rabbit polyclonal antibodies specific for S-SCAM (Hirao, K. et al. (1998) *J. Biol. Chem.* 273:21105-21110).

ISIGP-2 activity may be demonstrated by measuring the binding of ISIGP-2 to radiolabeled polypeptides containing the proline-rich region that specifically binds to WW containing proteins (Chen, H.I., and Sudol, M. (1995) *Proc. Natl. Acad. Sci. USA* 92:7819-7823). Samples of ISIGP-2 are run on SDS-PAGE gels, and transferred onto nitrocellulose by electroblotting. The blots are blocked for 1 hr at room temperature in TBST (137 mM NaCl, 2.7 mM KCl, 25 mM Tris (pH 8.0) and 0.1% Tween-20) containing non-fat dry milk. Blots are then incubated with TBST containing the radioactive formin polypeptide for 4 hrs to overnight. After washing the blots four times with TBST, the blots are exposed to autoradiographic film. Radioactivity is quantified by cutting out the radioactive spots and counting them in a radioisotope counter. The amount of radioactivity recovered is proportional to the activity of ISIGP-2 in the assay.

ISIGP-3 activity may be demonstrated by measuring the ability of ISIGP-3 to induce apoptosis. Mammalian cells (e.g., MCF7, HeLa, or NIH3T3 cells) are transfected with with 2 µg of

plasmid expressing ISIGP-3, or a control plasmid, together with 0.5  $\mu$ g of pCMV- $\beta$ -Gal using LipofectAMINE (Gibco-BRL). At defined times after transfection,  $\beta$ -Gal positive cells are counted and scored for characteristics of apoptosis, such as nuclear condensation and a shrunken, rounded morphology (Chan, S.-L. et al. (1999). J. Biol. Chem. 274:32461-32468). The percentage of  $\beta$ -Gal  
5 positive cells with an apoptotic morphology in ISIGP-3 transfected cells as compared to control cells is proportional to ISIGP-3 activity.

ISIGP-4 or ISIGP-5 activity may be demonstrated by measuring the ability of ISIGP to stimulate transcription of a reporter gene (Liu, H.Y. et al. (1997) EMBO J. 16:5289-5298). The assay entails the use of a well characterized reporter gene construct, LexA<sub>op</sub>-LacZ, that consists of LexA  
10 DNA transcriptional control elements (LexA<sub>op</sub>) fused to sequences encoding the *E. coli* LacZ enzyme. The methods for constructing and expressing fusion genes, introducing them into cells, and measuring LacZ enzyme activity, are well known to those skilled in the art. Sequences encoding ISIGP are cloned into a plasmid that directs the synthesis of a fusion protein, LexA-ISIGP, consisting of ISIGP and a  
DNA binding domain derived from the LexA transcription factor. The resulting plasmid, encoding a  
15 LexA-ISIGP fusion protein, is introduced into yeast cells along with a plasmid containing the LexA<sub>op</sub>-LacZ reporter gene. The amount of LacZ enzyme activity associated with LexA-ISIGP transfected cells, relative to control cells, is proportional to the amount of transcription stimulated by the ISIGP.

Various modifications and variations of the described methods and systems of the invention will  
20 be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following  
25 claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1309114	1	1309114CD1	6	1309114CBI
1478005	2	1478005CD1	7	1478005CBI
1597325	3	1597325CD1	8	1597325CBI
2791668	4	2791668CD1	9	2791668CBI
3223311	5	3223311CD1	10	3223311CBI



Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	1309114CD1	g10764778	0	phosphoinositol 3-phosphate-binding protein-2 [Homo sapiens] Dowler, S. et al. (2000) Identification of pleckstrin-homology-domain-containing proteins with novel phosphoinositide-binding specificities. Biochem. J. 351:19-31
2	1478005CD1	g4151807	5.0e-13	membrane-associated guanylate kinase-interacting protein Maguin-2 Yao, I. et al. (1999) J. Biol. Chem. 274:11889-11896
3	1597325CD1	g3930525	0.0	WW domain binding protein-1 [Homo sapiens] Chen, H.I. and Sudol, M. (1995) The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology3-binding modules. Proc. Natl. Acad. Sci. U.S.A. 92:7819-7823
4	2791668CD1	g6175860	5.0e-60	sex-determination protein homolog Fem1a [Mus musculus] Ventura-Holman, T. et al. (1998) The murine Fem1 gene family: Homologs of the Caenorhabditis elegans sex-determination protein FEM-1. Genomics 54:221-230
5	3223311CD1	g11611469	0.0	gl-related zinc finger protein [Mus musculus] Baker, S.J. and Reddy, E.P. (2000) Cloning of murine G1RP, a novel gene related to Drosophila melanogaster gl. Gene 248:33-40 deltex 2 [Mus musculus] Kishi, N. et al. (2001) Murine homologs of deltex define a novel gene family involved in vertebrate Notch signaling and neurogenesis. Int. J. Dev. Neurosci. 19:21-35

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1309114CD1	589	Y327 S55 T83 S494 T145 S149 S161 T252 S343 S445 S494 S526 T561 T60 T110 T225 S314 S351 S372 S382 T399 T467 S478	N129 N370 N380 N532	WW Domain 1: WL6-P41 WW domain: L12-P41, L58-P87 Pleckstrin homology domain: V170-L268 WW domain signature PR00403: L58-A71, Y73-P87	MOTIFS HMMEP_PPFAM HMMEP_PPFAM BLIMPS_PRINTS
2	1478005CD1	342	S168 S173 S250 S302 T247 T271 T326	N19 N89	Signal peptide: M1-H63 WW DOMAIN BINDING PROTEIN 1 PD129563: K11-S123; L133-M194; R242-E268	SPSCAN BLAST_PRODOR
3	1597325CD1	617	S145 S179 S240 S27 S278 S304 S333 S36 S370 S385 S418 T239 T242 T254 T407 T468 T563 Y167	N110 N111 N493 N582	Ankyrin repeats: N40-S73; E82-L114; T115-R147; H148-V180; K181-Y213; N481-S526; D527-L559 Ankyrin repeat signature PF00023: L45-L60 Ankyrin repeat PD00078: D525-L537 SEX DETERMINING PROTEIN FEM1 DEVELOPMENTAL PHOSPHORYLATION ANK REPEAT PD141329: K210-D396; P506-H616	HMMEP_PPFAM BLIMPS_PPFAM BLIMPS_PRODOR BLAST_PRODOR
4	2791668CD1	428	S10 S230 S241 S351 S352 S383 S50 S75 T127 T151 T262 T302 T410 T61	N101 N357 N385 N48 N59	Signal peptide: M1-A41 Transmembrane domain: Y206-Y229 Zinc finger, C3HC4 type (RING finger) domain: C277-C317 Zinc finger, C3HC4 type, signature: D273-E328 PHD-finger PF00628: I292-P306 ZINC FINGER, C3HC4 TYPE, DM00063 Q06003 119-171: D273-K323	HMMEP; SPSCAN HMMEP HMMEP_PPFAM PROFILESSCAN BLIMPS_PPFAM BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	3223311CD1	405	S156 S164 S254 T169 T312 T352 T368	N182 N373 N45	Zinc finger, C3HC4 type (RING finger) domain: C195-C255 PHD-finger PF00628: L225-A239 TRANSCRIPT DELTEX FRACTIONATED X- IRRADIATION INDUCED FX INDUCED THYMOMA CYTOPLASMIC BASIC PROTEIN PD021734: G298-E399	HMER_PFAM  BLIMPS_PFAM  BLAST_PRODUM

Table 4

Polynucleotide SEQ ID NO:// Incyte Polynucleotide ID/ Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
6 1309114CB1 2038	1-304, 960-1130, 2008-2038	g6140473	1456	2008
		1617090F6 (BRAITUT12)	1164	1742
		g1157664	1189	1882
		7264280H1 (PROSTMC02)	685	1380
		7374163H1 (ESOGTUE01)	1	561
		6243110H1 (TESTNOT17)	1584	1881
		g4690863	1650	2038
		7625427J1 (KIDNFEE02)	410	944
7 1478005CB1 2976	382-502, 1374-1511, 2205-2229, 2669-2699, 2846-2976	7738780H1 (BRAITUE01)	690	1325
		70794548V1	1383	2064
		573029R7 (BRAVUNT01)	467	1033
		6550830H1 (BRAFNON02)	1879	2534
		6839560H1 (BRSTNON02)	2049	2655
		6775282H1 (OVARDIR01)	1242	1827
		70772225V1	2355	2976
		5630841F6 (PLACFER01)	1	510
8 1597325CB1 2471	1-34, 485-619, 1168-1199, 2394-2471	921657T6 (RATRNOT02)	1794	2384
		71008458V1	1281	1835
		6264627H1 (MCLDTXN03)	1	581
		70845283V1	1914	2471
		71010679V1	675	1335
		71233907V1	1361	1884
		70469860V1	349	936
		4156408F6 (ADREN014)	1143	1611
9 2791668CB1 2796	1-24, 702-1125, 2037-2057, 2134-2796	1420994F6 (KIDNNOT09)	1899	2381
		2658667H1 (LUNGUT09)	1576	1829
		1365975R6 (SCORNON02)	2550	2796
		756115R1 (BRAITUT02)	2054	2584
		4733091H1 (SINTNOT19)	1385	1630
		6828289H1 (SINTNOR01)	433	1107
		6609076H2 (PLACFEC01)	3	564
		2771444H1 (COLANOT02)	1714	1954
10 3223311CB1 1992	1-678, 1510-1544, 1766-1992	6828289J1 (SINTNOR01)	576	1294
		456290F1 (KERANOT01)	1102	1758
		3775113H1 (BRSTNOT27)	1045	1340
		70515482V1	713	1300
		3464064F6 (293TF2T01)	199	644
		7315475H1 (SYNODIN02)	1	569
		2491754H1 (EOSITXT01)	577	817
		6433187H1 (LUNGNON07)	1489	1761
		2525302H1 (BRAITUT21)	1860	1992
		6117588H1 (SINITMT04)	1673	1965

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
6	1309114CB1	COLNFET02
7	1478005CB1	PLACFEB01
8	1597325CB1	PLACFER01
9	2791668CB1	BRAITUT02
10	3223311CB1	KERANOT01

Table 6

Library	Vector	Library Description
BRAITUT02	PSPORT1	Library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningioma lesion. Pathology indicated a grade 2 metastatic hypernephroma. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Family history included a malignant neoplasm of the kidney.
COLNFET02	PINCY	Library was constructed using RNA isolated from the colon tissue of a Caucasian female fetus, who died at 20 weeks' gestation.
KERANOT01	PLUESCRIPT	Library was constructed using RNA isolated from neonatal keratinocytes obtained from the leg skin of a spontaneously aborted black male.
PLACFEB01	PINCY	Library was constructed using pooled cDNA from two different donors. cDNA was generated using RNA isolated from placenta tissue removed from a Caucasian fetus (donor A), who died after 16 weeks' gestation from fetal demise and hydrocephalus; and a Caucasian male fetus (donor B), who died after 18 weeks' gestation from fetal demise. Patient history included umbilical cord wrapped around the head (3 times) and the shoulders (1 time) in donor A. Serology was positive for anti-CMV in donor A. Family history included multiple pregnancies and live births, and an abortion in donor A.
PLACFER01	PINCY	The library was constructed using RNA isolated from placental tissue removed from a Caucasian fetus, who died after 16 weeks' gestation from fetal demise and hydrocephalus. Patient history included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV. Family history included multiple pregnancies and live births, and an abortion.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDP	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A Blocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Artwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score ≥ GCG-specified 'HIGH' value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Cmsed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPSscan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	



What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ  
5 ID NO:1-5,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical  
to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5,
  - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from  
the group consisting of SEQ ID NO:1-5, and
  - 10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the  
group consisting of SEQ ID NO:1-5.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-5.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:6-  
20 10.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a  
polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method for producing a polypeptide of claim 1, the method comprising:
  - 30 a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell  
is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a  
promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
  - b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide selected from the group consisting of:
- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of
  - 5 SEQ ID NO:6-10,
  - b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10,
  - c) a polynucleotide complementary to a polynucleotide of a),
  - d) a polynucleotide complementary to a polynucleotide of b), and
  - 10 e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.
- 15 13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex
  - 20 is formed between said probe and said target polynucleotide or fragments thereof, and
  - b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.
- 25 15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:
- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
  - 30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.
16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

18. A method for treating a disease or condition associated with decreased expression of functional ISIGP, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional ISIGP, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional ISIGP, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of ISIGP in a biological sample comprising the steps of:

- a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e). a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of ISIGP in a subject, comprising administering to said subject an effective amount of the composition of claim 31.

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of ISIGP in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
- b) isolating antibodies from said animal; and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the

group consisting of SEQ ID NO:1-5.

36. An antibody produced by a method of claim 35.

5 37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

- 10 a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
- b) isolating antibody producing cells from the animal;
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;
- 15 d) culturing the hybridoma cells; and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

39. A monoclonal antibody produced by a method of claim 38.

20

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab expression library.

25

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 in a sample, comprising the steps of:

30

- a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide

having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5 from a sample, the method comprising:

5 a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

10 45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

15

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

20

50. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:6.

51. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:7.

52. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:8.

25

53. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:9.

54. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:10.

1/11



Asp	Leu	Cys	Leu	Phe	Tyr	Tyr	Arg	Asp	Glu	Lys	Glu	Glu	Gly	Ile
				200					205					210
Leu	Gly	Ser	Ile	Leu	Leu	Pro	Ser	Phe	Gln	Ile	Ala	Leu	Leu	Thr
				215					220					225
Ser	Glu	Asp	His	Ile	Asn	Arg	Lys	Tyr	Ala	Phe	Lys	Ala	Ala	His
				230					235					240
Pro	Asn	Met	Arg	Thr	Tyr	Tyr	Phe	Cys	Thr	Asp	Thr	Gly	Lys	Glu
				245					250					255
Met	Glu	Leu	Trp	Met	Lys	Ala	Met	Leu	Asp	Ala	Ala	Leu	Val	Gln
				260					265					270
Thr	Glu	Pro	Val	Lys	Arg	Val	Asp	Lys	Ile	Thr	Ser	Glu	Asn	Ala
				275					280					285
Pro	Thr	Lys	Glu	Thr	Asn	Asn	Ile	Pro	Asn	His	Arg	Val	Leu	Ile
				290					295					300
Lys	Pro	Glu	Ile	Gln	Asn	Asn	Gln	Lys	Asn	Lys	Glu	Met	Ser	Lys
				305					310					315
Ile	Glu	Glu	Lys	Lys	Ala	Leu	Glu	Ala	Glu	Lys	Tyr	Gly	Phe	Gln
				320					325					330
Lys	Asp	Gly	Gln	Asp	Arg	Pro	Leu	Thr	Lys	Ile	Asn	Ser	Val	Lys
				335					340					345
Leu	Asn	Ser	Leu	Pro	Ser	Glu	Tyr	Glu	Ser	Gly	Ser	Ala	Cys	Pro
				350					355					360
Ala	Gln	Thr	Val	His	Tyr	Arg	Pro	Ile	Asn	Leu	Ser	Ser	Ser	Glu
				365					370					375
Asn	Lys	Ile	Val	Asn	Val	Ser	Leu	Ala	Asp	Leu	Arg	Gly	Gly	Asn
				380					385					390
Arg	Pro	Asn	Thr	Gly	Pro	Leu	Tyr	Thr	Glu	Ala	Asp	Arg	Val	Ile
				395					400					405
Gln	Arg	Thr	Asn	Ser	Met	Gln	Gln	Leu	Glu	Gln	Trp	Ile	Lys	Ile
				410					415					420
Gln	Lys	Gly	Arg	Gly	His	Glu	Glu	Glu	Thr	Arg	Gly	Val	Ile	Ser
				425					430					435
Tyr	Gln	Thr	Leu	Pro	Arg	Asn	Met	Pro	Ser	His	Arg	Ala	Gln	Ile
				440					445					450
Met	Ala	Arg	Tyr	Pro	Glu	Gly	Tyr	Arg	Thr	Leu	Pro	Arg	Asn	Ser
				455					460					465
Lys	Thr	Arg	Pro	Glu	Ser	Ile	Cys	Ser	Val	Thr	Pro	Ser	Thr	His
				470					475					480
Asp	Lys	Thr	Leu	Gly	Pro	Gly	Ala	Glu	Glu	Lys	Arg	Arg	Ser	Met
				485					490					495
Arg	Asp	Asp	Thr	Met	Trp	Gln	Leu	Tyr	Glu	Trp	Gln	Gln	Arg	Gln
				500					505					510
Phe	Tyr	Asn	Lys	Gln	Ser	Thr	Leu	Pro	Arg	His	Ser	Thr	Leu	Ser
				515					520					525
Ser	Pro	Lys	Thr	Met	Val	Asn	Ile	Ser	Asp	Gln	Thr	Met	His	Ser
				530					535					540
Ile	Pro	Thr	Ser	Pro	Ser	His	Gly	Ser	Ile	Ala	Ala	Tyr	Gln	Gly
				545					550					555
Tyr	Ser	Pro	Gln	Arg	Thr	Tyr	Arg	Ser	Glu	Val	Ser	Ser	Pro	Ile
				560					565					570
Gln	Arg	Gly	Asp	Val	Thr	Ile	Asp	Arg	Arg	His	Arg	Ala	His	His
				575					580					585
Pro	Lys	Val	Lys											

<210> 2  
 <211> 342  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1478005CD1

identical  
 to SEQ ID  
 NO: 2

&lt;400&gt; 2

```

Met Pro Phe Leu Leu Gly Leu Arg Gln Asp Lys Glu Ala Cys Val
 1          5          10          15
Gly Thr Asn Asn Gln Ser Tyr Ile Cys Asp Thr Gly His Cys Cys
          20          25          30
Gly Gln Ser Gln Cys Cys Asn Tyr Tyr Glu Leu Trp Trp Phe
          35          40          45
Trp Leu Val Trp Thr Ile Ile Ile Ile Leu Ser Cys Cys Cys Val
          50          55          60
Cys His His Arg Arg Ala Lys His Arg Leu Gln Ala Gln Gln Arg
          65          70          75
Gln His Glu Ile Asn Leu Ile Ala Tyr Arg Glu Ala His Asn Tyr
          80          85          90
Ser Ala Leu Pro Phe Tyr Phe Arg Phe Leu Pro Asn Tyr Leu Leu
          95          100          105
Pro Pro Tyr Glu Glu Val Val Asn Arg Pro Pro Thr Pro Pro Pro
          110          115          120
Pro Tyr Ser Ala Phe Gln Leu Gln Gln Gln Gln Leu Leu Pro Pro
          125          130          135
Gln Cys Gly Pro Ala Gly Gly Ser Pro Pro Gly Ile Asp Pro Thr
          140          145          150
Arg Gly Ser Gln Gly Ala Gln Ser Ser Pro Leu Ser Glu Pro Ser
          155          160          165
Arg Ser Ser Thr Arg Pro Pro Ser Ile Ala Asp Pro Asp Pro Ser
          170          175          180
Asp Leu Pro Val Asp Arg Ala Ala Thr Lys Ala Pro Gly Met Glu
          185          190          195
Pro Ser Gly Ser Val Ala Gly Leu Gly Glu Leu Asp Pro Gly Ala
          200          205          210
Phe Leu Asp Lys Asp Ala Glu Cys Arg Glu Glu Leu Leu Lys Asp
          215          220          225
Asp Ser Ser Glu His Gly Ala Pro Asp Ser Lys Glu Lys Thr Pro
          230          235          240
Gly Arg His Arg Arg Phe Thr Gly Asp Ser Gly Ile Glu Val Cys
          245          250          255
Val Cys Asn Arg Gly His His Asp Asp Asp Leu Lys Glu Phe Asn
          260          265          270
Thr Leu Ile Asp Asp Ala Leu Asp Gly Pro Leu Asp Phe Cys Asp
          275          280          285
Ser Cys His Val Arg Pro Pro Gly Asp Glu Glu Glu Gly Leu Cys
          290          295          300
Gln Ser Ser Glu Glu Gln Ala Arg Glu Pro Gly His Pro His Leu
          305          310          315
Pro Arg Pro Pro Ala Cys Leu Leu Leu Asn Thr Ile Asn Glu Gln
          320          325          330
Asp Ser Pro Asn Ser Gln Ser Ser Ser Ser Pro Ser
          335          340

```

&lt;210&gt; 3

&lt;211&gt; 617

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1597325CD1

&lt;400&gt; 3

```

Met Asp Leu Lys Thr Ala Val Phe Asn Ala Ala Arg Asp Gly Lys
 1          5          10          15
Leu Arg Leu Leu Thr Lys Leu Leu Ala Ser Lys Ser Lys Glu Glu
          20          25          30
Val Ser Ser Leu Ile Ser Glu Lys Thr Asn Gly Ala Thr Pro Leu

```

identical  
to SEQ ID  
NO: 2

	35		40		45
Leu Met Ala Ala Arg Tyr Gly His Leu Asp Met Val Glu Phe Leu					
	50		55		60
Leu Glu Gln Cys Ser Ala Ser Ile Glu Val Gly Gly Ser Val Asn					
	65		70		75
Phe Asp Gly Glu Thr Ile Glu Gly Ala Pro Pro Leu Trp Ala Ala					
	80		85		90
Ser Ala Ala Gly His Leu Lys Val Val Gln Ser Leu Leu Asn His					
	95		100		105
Gly Ala Ser Val Asn Asn Thr Thr Leu Thr Asn Ser Thr Pro Leu					
	110		115		120
Arg Ala Ala Cys Phe Asp Gly His Leu Glu Ile Val Lys Tyr Leu					
	125		130		135
Val Glu His Lys Ala Asp Leu Glu Val Ser Asn Arg His Gly His					
	140		145		150
Thr Cys Leu Met Ile Ser Cys Tyr Lys Gly His Lys Glu Ile Ala					
	155		160		165
Gln Tyr Leu Leu Glu Lys Gly Ala Asp Val Asn Arg Lys Ser Val					
	170		175		180
Lys Gly Asn Thr Ala Leu His Asp Cys Ala Glu Ser Gly Ser Leu					
	185		190		195
Asp Ile Met Lys Met Leu Leu Met Tyr Cys Ala Lys Met Glu Lys					
	200		205		210
Asp Gly Tyr Gly Met Thr Pro Leu Leu Ser Ala Ser Val Thr Gly					
	215		220		225
His Thr Asn Ile Val Asp Phe Leu Thr His His Ala Gln Thr Ser					
	230		235		240
Lys Thr Glu Arg Ile Asn Ala Leu Glu Leu Leu Gly Ala Thr Phe					
	245		250		255
Val Asp Lys Lys Arg Asp Leu Leu Gly Ala Leu Lys Tyr Trp Lys					
	260		265		270
Lys Ala Met Asn Met Arg Tyr Ser Asp Arg Thr Asn Ile Ile Ser					
	275		280		285
Lys Pro Val Pro Gln Thr Leu Ile Met Ala Tyr Asp Tyr Ala Lys					
	290		295		300
Glu Val Asn Ser Ala Glu Glu Leu Glu Gly Leu Ile Ala Asp Pro					
	305		310		315
Asp Glu Met Arg Met Gln Ala Leu Leu Ile Arg Glu Arg Ile Leu					
	320		325		330
Gly Pro Ser His Pro Asp Thr Ser Tyr Tyr Ile Arg Tyr Arg Gly					
	335		340		345
Ala Val Tyr Ala Asp Ser Gly Asn Phe Lys Arg Cys Ile Asn Leu					
	350		355		360
Trp Lys Tyr Ala Leu Asp Met Gln Gln Ser Asn Leu Asp Pro Leu					
	365		370		375
Ser Pro Met Thr Ala Ser Ser Leu Leu Ser Phe Ala Glu Leu Phe					
	380		385		390
Ser Phe Met Leu Gln Asp Arg Ala Lys Gly Leu Leu Gly Thr Thr					
	395		400		405
Val Thr Phe Asp Asp Leu Met Gly Ile Leu Cys Lys Ser Val Leu					
	410		415		420
Glu Ile Glu Arg Ala Ile Lys Gln Thr Gln Cys Pro Ala Asp Pro					
	425		430		435
Leu Gln Leu Asn Lys Ala Leu Ser Ile Ile Leu His Leu Ile Cys					
	440		445		450
Leu Leu Glu Lys Val Pro Cys Thr Leu Glu Gln Asp His Phe Lys					
	455		460		465
Lys Gln Thr Ile Tyr Arg Phe Leu Lys Leu His Pro Arg Gly Lys					
	470		475		480
Asn Asn Phe Ser Pro Leu His Leu Ala Val Asp Lys Asn Thr Thr					
	485		490		495
Cys Val Gly Arg Tyr Pro Val Cys Lys Phe Pro Ser Leu Gln Val					
	500		505		510

```

Thr Ala Ile Leu Ile Glu Cys Gly Ala Asp Val Asn Val Arg Asp
515 520 525
Ser Asp Asp Asn Ser Pro Leu His Ile Ala Ala Leu Asn Asn His
530 535 540
Pro Asp Ile Met Asn Leu Leu Ile Lys Ser Gly Ala His Phe Asp
545 550 555
Ala Thr Asn Leu His Lys Gln Thr Ala Ser Asp Leu Leu Asp Glu
560 565 570
Lys Glu Ile Ala Lys Asn Leu Ile Gln Pro Ile Asn His Thr Thr
575 580 585
Leu Gln Cys Leu Ala Ala Arg Val Ile Val Asn His Arg Ile Tyr
590 595 600
Tyr Lys Gly His Ile Pro Glu Lys Leu Glu Thr Phe Val Ser Leu
605 610 615
His Arg

```

&lt;210&gt; 4

&lt;211&gt; 428

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2791668CD1

&lt;400&gt; 4

```

Met Gly Pro Pro Pro Gly Ala Gly Val Ser Cys Arg Gly Gly Cys
1 5 10 15
Gly Phe Ser Arg Leu Leu Ala Trp Cys Phe Leu Leu Ala Leu Ser
20 25 30
Pro Gln Ala Pro Gly Ser Arg Gly Ala Glu Ala Val Trp Thr Ala
35 40 45
Tyr Leu Asn Val Ser Trp Arg Val Pro His Thr Gly Val Asn Arg
50 55 60
Thr Val Trp Glu Leu Ser Glu Glu Gly Val Tyr Gly Gln Asp Ser
65 70 75
Pro Leu Glu Pro Val Ala Gly Val Leu Val Pro Pro Asp Gly Pro
80 85 90
Gly Ala Leu Asn Ala Cys Asn Pro His Thr Asn Phe Thr Val Pro
95 100 105
Thr Val Trp Gly Ser Thr Val Gln Val Ser Trp Leu Ala Leu Ile
110 115 120
Gln Arg Gly Gly Gly Cys Thr Phe Ala Asp Lys Ile His Leu Ala
125 130 135
Tyr Glu Arg Gly Ala Ser Gly Ala Val Ile Phe Asn Phe Pro Gly
140 145 150
Thr Arg Asn Glu Val Ile Pro Met Ser His Pro Gly Ala Val Asp
155 160 165
Ile Val Ala Ile Met Ile Gly Asn Leu Lys Gly Thr Lys Ile Leu
170 175 180
Gln Ser Ile Gln Arg Gly Ile Gln Val Thr Met Val Ile Glu Val
185 190 195
Gly Lys Lys His Gly Pro Trp Val Asn His Tyr Ser Ile Phe Phe
200 205 210
Val Ser Val Ser Phe Phe Ile Ile Thr Ala Ala Thr Val Gly Tyr
215 220 225
Phe Ile Phe Tyr Ser Ala Arg Arg Leu Arg Asn Ala Arg Ala Gln
230 235 240
Ser Arg Lys Gln Arg Gln Leu Lys Ala Asp Ala Lys Lys Ala Ile
245 250 255
Gly Arg Leu Gln Leu Arg Thr Leu Lys Gln Gly Asp Lys Glu Ile
260 265 270

```

Gly	Pro	Asp	Gly	Asp	Ser	Cys	Ala	Val	Cys	Ile	Glu	Leu	Tyr	Lys
				275					280					285
Pro	Asn	Asp	Leu	Val	Arg	Ile	Leu	Thr	Cys	Asn	His	Ile	Phe	His
				290					295					300
Lys	Thr	Cys	Val	Asp	Pro	Trp	Leu	Leu	Glu	His	Arg	Thr	Cys	Pro
				305					310					315
Met	Cys	Lys	Cys	Asp	Ile	Leu	Lys	Ala	Leu	Gly	Ile	Glu	Val	Asp
				320					325					330
Val	Glu	Asp	Gly	Ser	Val	Ser	Leu	Gln	Val	Pro	Val	Ser	Asn	Glu
				335					340					345
Ile	Ser	Asn	Ser	Ala	Ser	Ser	His	Glu	Glu	Asp	Asn	Arg	Ser	Glu
				350					355					360
Thr	Ala	Ser	Ser	Gly	Tyr	Ala	Ser	Val	Gln	Gly	Thr	Asp	Glu	Pro
				365					370					375
Pro	Leu	Glu	Glu	His	Val	Gln	Ser	Thr	Asn	Glu	Ser	Leu	Gln	Leu
				380					385					390
Val	Asn	His	Glu	Ala	Asn	Ser	Val	Ala	Val	Asp	Val	Ile	Pro	His
				395					400					405
Val	Asp	Asn	Pro	Thr	Phe	Glu	Glu	Asp	Glu	Thr	Pro	Asn	Gln	Glu
				410					415					420
Thr	Ala	Val	Arg	Glu	Ile	Lys	Ser							
				425										

&lt;210&gt; 5

&lt;211&gt; 405

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3223311CD1

&lt;400&gt; 5

Met	Thr	Asn	Leu	Pro	Ala	Tyr	Pro	Val	Pro	Gln	His	Pro	Pro	His
1				5					10					15
Arg	Thr	Ala	Ser	Val	Phe	Gly	Thr	His	Gln	Ala	Phe	Ala	Pro	Tyr
				20					25					30
Asn	Lys	Pro	Ser	Leu	Ser	Gly	Ala	Arg	Ser	Ala	Pro	Arg	Leu	Asn
				35					40					45
Thr	Thr	Asn	Ala	Trp	Gly	Ala	Ala	Pro	Pro	Ser	Leu	Gly	Ser	Gln
				50					55					60
Pro	Leu	Tyr	Arg	Ser	Ser	Leu	Ser	His	Leu	Gly	Pro	Gln	His	Leu
				65					70					75
Pro	Pro	Gly	Ser	Ser	Thr	Ser	Gly	Ala	Val	Ser	Ala	Ser	Leu	Pro
				80					85					90
Ser	Gly	Pro	Ser	Ser	Ser	Pro	Gly	Ser	Val	Pro	Ala	Thr	Val	Pro
				95					100					105
Met	Gln	Met	Pro	Lys	Pro	Ser	Arg	Val	Gln	Gln	Ala	Leu	Ala	Gly
				110					115					120
Met	Thr	Ser	Val	Leu	Met	Ser	Ala	Ile	Gly	Leu	Pro	Val	Cys	Leu
				125					130					135
Ser	Arg	Ala	Pro	Gln	Pro	Thr	Ser	Pro	Pro	Ala	Ser	Arg	Leu	Ala
				140					145					150
Ser	Lys	Ser	His	Gly	Ser	Val	Lys	Arg	Leu	Arg	Lys	Met	Ser	Val
				155					160					165
Lys	Glu	Ala	Thr	Pro	Lys	Pro	Glu	Pro	Glu	Pro	Glu	Gln	Val	Ile
				170					175					180
Lys	Asn	Tyr	Thr	Glu	Glu	Leu	Lys	Val	Pro	Pro	Asp	Glu	Asp	Cys
				185					190					195
Ile	Ile	Cys	Met	Glu	Lys	Leu	Ser	Ala	Ala	Ser	Gly	Tyr	Ser	Asp
				200					205					210
Val	Thr	Asp	Ser	Lys	Ala	Ile	Gly	Ser	Leu	Ala	Val	Gly	His	Leu
				215					220					225

Thr	Lys	Cys	Ser	His	Ala	Phe	His	Leu	Leu	Cys	Leu	Leu	Ala	Met
				230						235				240
Tyr	Cys	Asn	Gly	Asn	Lys	Asp	Gly	Ser	Leu	Gln	Cys	Pro	Ser	Cys
				245						250				255
Lys	Thr	Ile	Tyr	Gly	Glu	Lys	Thr	Gly	Thr	Gln	Pro	Gln	Gly	Lys
				260						265				270
Met	Glu	Val	Leu	Arg	Phe	Gln	Met	Ser	Leu	Pro	Gly	His	Glu	Asp
				275						280				285
Cys	Gly	Thr	Ile	Leu	Ile	Val	Tyr	Ser	Ile	Pro	His	Gly	Ile	Gln
				290						295				300
Gly	Pro	Glu	His	Pro	Asn	Pro	Gly	Lys	Pro	Phe	Thr	Ala	Arg	Gly
				305						310				315
Phe	Pro	Arg	Gln	Cys	Tyr	Leu	Pro	Asp	Asn	Ala	Gln	Gly	Arg	Lys
				320						325				330
Val	Leu	Glu	Leu	Leu	Lys	Val	Ala	Trp	Lys	Arg	Arg	Leu	Ile	Phe
				335						340				345
Thr	Val	Gly	Thr	Ser	Ser	Thr	Thr	Gly	Glu	Thr	Asp	Thr	Val	Val
				350						355				360
Trp	Asn	Glu	Ile	His	His	Lys	Thr	Glu	Met	Asp	Arg	Asn	Ile	Thr
				365						370				375
Gly	His	Gly	Tyr	Pro	Asp	Pro	Asn	Tyr	Leu	Gln	Asn	Val	Leu	Ala
				380						385				390
Glu	Leu	Ala	Ala	Gln	Gly	Val	Thr	Glu	Asp	Cys	Leu	Glu	Gln	Gln
				395						400				405

&lt;210&gt; 6

&lt;211&gt; 2038

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1309114CB1

&lt;400&gt; 6

```

gcgcgcgcggg ccgggggaggc gcgcctcgtc cgcgcctccct tcgcctcgtc gtttccctcct 60
ccctcggcag ccgcggcggc agcaggagaa ggcgcgcgcg gcggctaggg atcagacatg 120
gcggcggaac tgaacctgga gtggatctcc ctgcccgggt cctggactta cgggatcacc 180
agggcgcgcc gagtcttctt catcaacgag gaggccaaga gcaccacctg gctgcacccc 240
gtcaccggcg aggcgggtgt caccggacac cggcggcaga gcacagattt gctactggc 300
tggaagaag catatacttt tgaagggtca agatactata taaaccataa tgaaaggaaa 360
gtgacctgca aacatccagt cacaggacaa ccatcacagg acaattgtat ttttgtagt 420
aatgaacaga ctgttgcaac catgacatct gaagaaaaga aggaacggcc aataagtatg 480
ataaatgaag cttctaacta taacgtgact tcagattatg cagtgcaccc aatgagccct 540
gtaggcagaa cttcacgagc ttcaaaaaaa gttcataatt ttggaaagag gtcaaattca 600
attaaaagga atcctaagtc accggttgtc agacgaggtt ggctttataa acaggacagt 660
actggcatga aattgtggaa gaaacgctgg tttgtgcttt ctgacctttg cctcttttat 720
tatagagatg agaaagaaga gggatctctg ggaagcatac tgttacctag ttttcagata 780
gctttgctta cctctgaaga tcacattaat cgcaaatatg cttttaaggc agcccatcca 840
aacatgcgga cctattattt ctgactgat acaggaaagg aaatggagtt tgggatgaaa 900
gccatgttag atgctgccct agtacagaca gaacctgtga aaagagtggg caagattaca 960
tctgaaaatg caccaactaa agaaaccaat aacattccca accatagagt gctaattaaa 1020
ccagagatcc aaaacaatca aaaaaacaag gaaatgagca aaattgaaga aaaaaaggca 1080
ttagaagctg aaaaatatgg atttcagaag gatggtcaag atagaccctt aacaaaaatt 1140
aatagtgtaa agctgaattc tctgccatct gaatatgaga gtgggtcagc atgcccgtgt 1200
cagactgtgc actacagacc aatcaacttg agcagttcag agaacaaaat agtcaatgtt 1260
agcctggcag atcttagagg tggaaatcgc ccaatacag ggcccttata cacagaggcc 1320
gatcgagtca tacagagaac aaattcaatg cagcagttgg aacagtggtt taaaatccag 1380
aaggggaggg gtcatgaaga agaaaccagg ggagtaattt cttaccaaac attaccaaga 1440
aatatgcaa gtcacagagc ccagattatg gcccgctacc ctgaagggtt tagaacactc 1500
ccaagaaaca gcaagacaag gcctgaaagt atctgcagt taaccccttc cactcagac 1560
aagacattag gaccgggagc ggaggagaaa cggaggtcca tgagagatga cacaatgttg 1620

```

```

cagctctacg aatggcagca gcgtcagttt tataacaaac agagcaccct ccctcgacac 1680
agtactttga gtagtcccaa aaccatggta aatattttctg accagacaat gcaactctatt 1740
cccacatcac cttccacagg gtcaatagct gcttatcagg gatactcccc tcaacgaact 1800
tacagatcgg aagtgtcttc accaattcag agaggagatg tgacaataga ccgcagacac 1860
agggcccatc accctaaggt aaaaatagctg ctgattttgt gttaactcac taccttataa 1920
atgctgtgtt ttctttctag tatactattt taaatgtgag agacaaaaga atggggataa 1980
agtaagcaag gcagctcttt tttgttttaa aaaataaata aaaatatattt acaacaaa 2038

```

&lt;210&gt; 7

&lt;211&gt; 2976

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1478005CB1

&lt;400&gt; 7

```

cttgatttat gtacccccca gcctgcttag agccaagggg ttgcagcagc ctgctcccat 60
ctgcagcccc caccatcctc ccacagtggg ctctggctct aggtgggtcc agggctgggc 120
atcgcggttg tgcagcacat cctcctcagt attccagtgc agctgtctga agttttttct 180
gctgcgcctg aactgatgtc atttccccct tggcagacag cttcggcttt gctgcgtctg 240
agatatgtca cgagaaggtg ggggtgggccc agagccaggc agggggagta gcgaggagag 300
caggagacag tgtgcctgct cgggtcccagg actctgttta ctttgtctgc tttgctaaag 360
aaggccggtg aaccaggacc accgcacaca caggcccacc aggggcaatg ctcattccaa 420
gaccttaact tttaagagcc ctttgttcca acgttagtgt ggacgatgct cttgcaggat 480
gcctttcctt ttgggtctta gacaggataa ggaagcctgt gtgggtacca acaatcaaag 540
ctacatctgt gacacaggac actgctgttg acagtctcag tgcgtcaact actactatga 600
actctgtgtg ttctggctgg tgtggaccat catcatcatc ctgagctgct gctgtgtttg 660
ccaccaccgc cgagccaagc accgccttca ggcccagcag cggaacatg aaatcaacct 720
gatectttac cgagaagccc acaattactc agcgttgcca ttttatttca ggtttttgccc 780
aaactattta ctacctcctt atgaggaagt ggtgaaccga cctccaactc ctccccacc 840
atacagtgcc ttccagctac agcagcagca gctgctgcct ccacagtgtg gccctgcagg 900
tggcagtcct ccgggcatcg atccccaccg gggatcccag ggggcacaga gcagccccct 960
gtctgagccc agcagaagca gcacaagacc cccaagcatc gctgaccctg atccctctga 1020
cctaccagtt gaccgagcag ccaccaagc ccagggatg gagcccagtg gctctgtggc 1080
tggcctgggg gagctggacc cgggggcctt cctggacaaa gatgcagaat gtagggagga 1140
gctgctgaaa gatgacagct ctgaacacgg cgcaccgac agcaaagaga agacgcctgg 1200
gagacatcgc cgcttcacag gtgactcggg cattgaagtg tgtgtgtgca accggggcca 1260
ccatgacgat gacctcaaag agttcaacac actcatcgat gatgctctgg atgggcccct 1320
ggacttctgc gacagctgcc atgtgcggcc ccttggtgat gagggaggaa gccctgtgca 1380
gtcctctgag gagcaggtc gagagcctgg gcaccgcac ctgccacggc cgcccgcag 1440
cctgctgctg aacaccatca acgagcagga ctctcccaac tcccagagca gcagctcccc 1500
cagctagagc aggtcctgcc agcaccagc aacttggcaa agcaaccagg gtaggggaga 1560
accacgagag aagcattaag tgactttcaa agactttcag agtacagcca cttgggttct 1620
ttttgtttgt ttctctctc ctctcctgca tttctctcca tctccaggta cagttcgggg 1680
tgtggatgcc tcttctctca caagggcaca gtgttgtgga gggctaagtt gggtctgtga 1740
ctcattcctc ataccctaac tccatctcct ttctttaaag tcaaactca cctacctgtt 1800
tgggtcagag agatgtgttt taaaagcccc caaggaagga ggctgggact gtgccctgac 1860
atgattcttg gtgatggaat aggtttgtgc tctgattcta gtttaagaga acgttgctgt 1920
atctcagtc aggagaggca gcccatcttg gccctggatg aagaaggaaa cccacagagg 1980
cccaggcctt gtcattgggc tgccagtgtc tgccaagcca gcattgagct aatcctgtgg 2040
gaggatgaga gctactgggc cgttgtatga taggttggta ggggcttgtt gatctgtcaa 2100
attccaggtg acaagatcta tgcaccccat gcgtccttga ggggcctctt ccccgaggc 2160
tctggctggc cgcaggctgg ttctggtgtg aaaggttata ctgccttttc tttgtttgtt 2220
tgttttttct tctaaaaaca aacagcaaaa gacagctgaa aacaagaact tcaccgggtg 2280
gcaggcaaga attctctctt ggaaatgac gtttgtggct ctttcccaag ttggccttca 2340
aagagcctgc ctgctgttga gccagaagat gtctcgtgtg aaggctgggg ttggcggctgt 2400
cttggaacct ctgtgagcag gaggccctaa gccgcagcag tggatagagg tgcagctctc 2460
tgctctctct ccttttggtc tgtgttcaca ggtgaccctg gtcagcctgc atcgcaagca 2520
cacaccctgc gggccttcaa gtctcactgt tccgtatgag gaaacagaca gcggactgag 2580
gaagcgatgg cccagagaaa agggcccttg tagcctggct ctacacagct attttatctt 2640
tgattctgaa taaatatattt ttgtgggggt tttttttttt ttgggggggc agtgggttgg 2700

```

```

tttaaaactg accactggga agaaacacct ggggttatcgg ggggtttccat gcctggtcct 2760
tgcccttttac ccccaaccct tttggagtcg ggtgccatt ttcctgtgta gagactcggg 2820
ggcccaggca ggaggtgaaa gcagcattcg gaaggccctg ggggaccctt ggggcttggtg 2880
gcccgccctt cgggtcacca gttgagctgc gatgggaaac tctgatgggc gcgcgcaacg 2940
gcaaaacaat ttttcccaac gggcttgtga tatgag 2976

```

&lt;210&gt; 8

&lt;211&gt; 2471

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1597325CB1

&lt;400&gt; 8

```

gcaggcggag caggcgggcc cgggcggcgg tggggacaac ggtttccctt tgaaggggac 60
ggacaaagcc cgagtgacca gcggcggcgg ggaggactag tccccgggca gtttggtgcc 120
ctgggtgtgca gatgttgga agcagtagga cggaacatac tcttcgtggg tgtgtatccg 180
ttctgggggtg cagcaattaa cattggactt tggttcctgt gactcttgcc tgtgtcgata 240
gagtttaaact ggagctctgc tttgaaagat aaataaagca cagcctctca actggacata 300
aatggatcta aagacagcag tatttaacgc agctcgggat ggcaaaactcc ggcttctcac 360
caaattgttg gcaagcaaat ccaaagagga ggtttccctc ttgatctctg aaaaaacaaa 420
tggggccacg ccactcttga tggccgccag gtatgggcac cttgacatgg tggaaattcc 480
cctagagcaa tgcagtgcct ccatagaagt tgggggctcc gtcaattttg atggcgaaac 540
cattgagggg gctccccctt tatgggccgc ttctgcagca ggacatctga aggtggtcca 600
gtccttggtta aatcatggag catctgtcaa caacacgact ttaaccaatt caactcctct 660
tcgagctgcg tgtttcgatg gccatttgga aatagtgaag taccttgtag aacacaaagc 720
tgatttgga gtgtcaaacc gacatgggca tacgtgcttg atgatttcat gttacaaagg 780
acataaagag attgctcagt atttacttga aaagggggca gatgttaata gaaaaagtgt 840
caaaggtaat actgcattgc atgattgtgc agaactctgga agtttggaca tcatgaagat 900
gcttcttatg tattgtgcca agatggaaaa ggaatggttat ggaatgactc ccttctctc 960
agcaagtgtg actggtcaca caaatattgt ggattttctg acacaccatg cacagaccag 1020
caagacagaa cgtattaatg ctctagagct tctgggagct acatttgtag acaaaaaaag 1080
agatctgctt ggggctttga aatactggaa aaaggcaatg aacatgaggt acagtgatag 1140
gactaatatt attagtaaac cagtgccaca gacactaata atggcttatg attatgcaa 1200
ggaagtgaac agtgcagaag agctagaagg tcttattgct gatcctgatg agatgagaat 1260
gcaggcacta ttaatcagag aacgtattct tggtecttct catcctgata cctcttacta 1320
tattagatat agagcgctg tctatgcaga ctctggaaat ttcaaacgat gcacaaacct 1380
atggaagtat gctttggata tgcagcagag caatttggat cctttaagcc caatgaccgc 1440
cagcagctta ttatcttttg cagaactatt ctctttatg ctacaggata gggctaaagg 1500
cctgctgggt actactgtta catttgatga tcttatgggc atactttgca aaagcgtcct 1560
tgaaatagag cgagctatca aacaaactca gtgtccagct gaccattac agttaaataa 1620
ggccctttct attattttgc acttaatttg cttgttagag aaagtccctt gtactctaga 1680
acaagaccat ttcaaaaagc agactatata caggtttctt aagctgcac caaggggaaa 1740
gaataacttc agccctcttc atctggctgt ggacaagaat actacatgtg tagggcggta 1800
ccctgtttgt aaatttccat ctctacaagt tactgcaata ctgatagaat gtggtgctga 1860
tgtgaacgtc agagactcgg atgacaacag tcccctgcat atcgctgctc ttaacaacca 1920
tccagacatc atgaatctcc ttattaaatc aggtgcacat tttgatgcca caaacttgca 1980
caaacaaact gctagtgact tgctggatga gaaggaaata gctaaaaatt taatccagcc 2040
tataaatcat accacattgc agtgtcttgc tgctcgtgtc atagtgaatc atagaatata 2100
ttataaaggg catatcccag aaaagctaga gacttttgtt tcccctcata gatgataact 2160
tgactgtatt ttagcactgt taaagcacga attggttaaca gttgtttcat aaatgagcac 2220
tgttgtgata acaccagcat tcatttagct tgattgatat cattgtgctc tcattggcta 2280
aagcattata agcatcaaat ttacaacatt ggtttcccaa tatttaatat aaatatacca 2340
tataatatat tgtttgtgaa ttattgagaa atgtaacatt caaatttcta aaattgtctg 2400
ccaaaggctt attcattctg gttttgtttg ctgttgggtg tttggggcag agttaacct 2460
ttctccatgg t 2471

```

&lt;210&gt; 9

&lt;211&gt; 2796

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens



&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2791668CB1

&lt;400&gt; 9

```
agcgcggtag cggagaagac tggagctccg aggagctgca tctgcggcaa cctgtgtgct 60
gacgctacgt gcctcctggc tccgacgtag ctgcagctc cccagtctca ctccattcct 120
tccccacctg gcgcgcacct gctcaagacc agggtcctgc caagcgctag gagggcgcgt 180
gccaggggag ctaggggaact gcggagcgcg cgcgccatgg ggcgcgcgcc tggggccggg 240
gtctcctgcc gcggtggotg cggcttttcc agattgctgg catggtgctt cctgctggcc 300
ctgagtcctc aggcacccgg ttcccggggg gctgaagcag tgtggaccgc gtacctcaac 360
gtgtcctggc ggggttccgca cacgggagtg aaccgtacgg tgtgggagct gacgaggag 420
ggcgtgtacg gccaggactc gccgctggag cctgtggctg gggctcctgg accgcccag 480
gggcccgggg cgcttaacgc ctgtaaccgg cacacgaatt tcacggtgcc cacggtttgg 540
ggaagcacct tgcaagtctc ttggttggcc ctcatccaac ggcggggggg ctgcaccttc 600
gcagacaaga tccatctggc ttatgagaga ggggcgtctg gagccgtcat ctttaacttc 660
cccgggaccc gcaatgaggt catcccatg tctcaccggg gtgcagtaga cattgttgca 720
atcatgatcg gcaatctgaa aggcacaaaa attctgcaat ctattcaaag aggcatacaa 780
gtgacaatgg tcatagaagt agggaaaaaa catggccctt gggtgaaatc ctattcaatt 840
tttttcgttt ctgtgtcctt ttttattatt acggcgccaa ctgtgggcta ttttatcttt 900
tattctgtct gaaggctacg gaatgcaaga gctcaaagca ggaagcagag gcaattaaag 960
gcagatgcta aaaaagctat tggaaaggct caactacgca cactgaaaca aggagacaag 1020
gaaattggcc ctgatggaga tagttgtgct gtgtgcattg aattgtataa accaaatgat 1080
ttggtagcga tcttaacgtg caaccatatt ttccataaga catgtgttga ccatggctg 1140
ttagaacaca ggacttgccc catgtgcaaa tgtgacatac tcaaagcttt gggaattgag 1200
tgggatggtt aagatggatc agtgtcctta caagtccctg tatccaatga aatatctaatt 1260
agtgcctcct cccatgaaga ggataatcgc agcgagaccg catcatctgg atagcttca 1320
gtacagggaa cagatgaacc gcctctggag gaacacgtgc agtcaacaaa tgaagtgcta 1380
cagctggtaa accatgaagc aaattctgtg gcagtggtat ttattcctca tgttgacaac 1440
ccaacctttg aagaagacga aactcctaatt caagagactg ctgttcgaga aattaaatct 1500
taaaaatctg gtaaatagaa aacttgaaac attagtaata acagaactgc caatcagggc 1560
ctagtttcta ttaataaatt ggataaattt aataaaataa gagtgtactt gaaagtgtct 1620
agatgactaa tattatgcta tagttaaatg gcttaaaata tttaacctgt taactttttt 1680
ccacaaactc attataatat ttttcatagg caagtttctc ctacgttagt ataacaacat 1740
tttttagacat tcaaaactgt cttcaagaag tcacgttttt catttataac aattttctta 1800
taaaaactgt ttgcttttaa aatgtggagt agctgtaac actttatttt atgatagtat 1860
cttaatgaaa aatactactt ctttagcttg ggctacatgt gtcagggttt ttctccaggt 1920
gcttatattg atctggaatt gtaatgtaaa aagcaatgca aacttaggcg agtacttctt 1980
gaaatgtcta tttaagctgc tttaagttaa tagaaaagat taaagcaaaa tattcatttt 2040
tactttttct tattttttaa attaggctga atgtacttca tgtgatttgt caaccatagt 2100
ttatcgagaga ttatggactt aattgattgg tatattagtg acatcaactt gacacaagat 2160
tagacaaaaa attccttaca aaaatactgt gtaactattt ctcaaacttg tgggattttt 2220
caaaagctca gtatatgaat catcactactg tttgaaattg ctaatgacag agtaagtaac 2280
actaatattg gtcattgatc ttcgttcatg aattagtcta cagaaaaaaa atgttctgta 2340
aaattagtct gttgaaaatg ttttccaaac aatgttactt tgaaaaattga gtttatgttt 2400
gacctaaatg ggctaaaatt acattagata acataaaatt ctgtccgtgt aactataaat 2460
tttgatgaat cattttcctg gtgtttgaaa aagaaggggg ggagaattcc aggtgcctta 2520
atataaagtt tgaagcttca tccaccaaag ttaaatagag ctatttataa atgcacttta 2580
tttgactctt gtgtggcttt tgttttagaa ttttgttcaa attatagcag aatttaggca 2640
aaaaataaac agacatgtat ttttgtttgc tgaatggatg aaaccattgc attcttgtac 2700
actgatttga aatgctgtaa atatgtccca atttgattg attctcttta aatataaaat 2760
gtaaataaaa tattccaata aaaaaaaaaa aaaaaa 2796
```

&lt;210&gt; 10

&lt;211&gt; 1992

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3223311CB1

&lt;400&gt; 10

```

agctttttgcc gcagcgtgcg ggcgcaagca gggccgcctt acccgggtgac caccatcatc 60
gtcccgccgg gccacacagg cgtcgctgc tcttgccacc agtgccctcag tggcagcaga 120
actggccccc tgtagggccg ctaccgccac tccatgacca acctccctgc ataccccgtc 180
ccccagcacc cccacacag gaccgcttct gtgtttggga cccaccaggc ctttgaccgc 240
tacaacaaac cctcactctc cggggcccg tctgcgccc ggctgaacac caccaacgcc 300
tggggcgag cctcctcttc cctggggagc cagccctct accgctccag cctctccac 360
ctgggaccgc agcacctgcc cccaggatcc tccacctccg gtgcagtcag tgcctccctc 420
cccagcggtc cctcaagcag cccaggggagc gtccctgcca ctgtgcccat gcagatgcca 480
aagcccgca gagtcagca ggcgctcgca ggcagtcag gtgttctgat gtcagccatt 540
ggactccctg tgtgtcttag ccgcgcaccc cagccacca gccctccgc cctccgctctg 600
gcttccaaaa gtcacggctc agttaagaga ttgaggaaaa tgcctgtaa agaagcgacc 660
ccgaagccag agccagagcc agagcaggtc ataaaaaact acacggaaga gctgaaagt 720
ccccagatg aggactgcat catctgcat gagaagctgt ccgcagcgtc tggatacagc 780
gatgtgactg acagcaaggc aatcgggtcc ctggctgtgg gccacctcac caagtgcagc 840
catgccttcc acctgtgtg cctcctggcc atgtactgca acggcaataa ggatggaagt 900
ctgcagtgtc cctcctgcaa aaccatctat ggagagaaga cggggaccca gcccaggga 960
aagatggagg tattacggtt ccagatgtcg cctcccgccc acgaggactg cgggaccatc 1020
ctcatagttt acagcattcc ccatggatc cagggccctg agcaccacca tcccggaaag 1080
ccgttcactg ccagagggtt tccccgccag tgctacctc cagacaacgc ccaggcgccg 1140
aaggtcctag agtcctgaa ggtggcctgg aagaggcgcc tcactctcac agtgggcacg 1200
tccagcacca cgggtgagac ggacaccgtg gtatggaacg agatccacca caagacagag 1260
atggaccgca acattacggg ccacggctat cccgaccca actacctgca gaacgtgctg 1320
gctgagctgg ctgcccaggg ggtgaccgag gactgcctgg agcagcagtg acctcgacc 1380
ccagcacgcc cgcctctggt ggccaccccg ctgcccctat gctggctggg tggccaggca 1440
ggaagtgcgc agcccagag gctgggaggt ttgttgaggg tgtggggtgt gccccacctg 1500
aagccggggc tccccctgcc tgcctctctc tctcctccc ctctgggaat tgggcagccc 1560
tgggcagttg tactcatggg ggcttaggat gcagctacct cagtgcgcag ggcccgctc 1620
tctctgggg gctgcttcgg gcccgcggtg ctggggcct ggtgtggggc gagtagagac 1680
ttccccagcc tggacggcg tgggttctgg gtcagcttct ttacctcaa tttgtttgc 1740
aataaatgct ctatagcaa agccagcagg tctgagtggt gtgcagtcag cgtgtgtgtc 1800
gcacttgtgt gtgtgtgtgc cccccccac ttctgcac agagcaagag ggggttccat 1860
gggctcatcg gctccattt gataactgaa gaacaggcca cagccaggca tggaggagcc 1920
cacggtactg ggctgtgcgg cctccacatg ccctacactg atctccctgc catgccagag 1980
gctgtcacc ca 1992

```